

Covariance fields

Nikolay Balov *

balov@stat.fsu.edu

January 15, 2009

Abstract

We introduce and study covariance fields of distributions on a Riemannian manifold. At each point on the manifold, covariance is defined to be a symmetric and positive definite (2,0)-tensor. Its product with the metric tensor specifies a linear operator on the respected tangent space. Collectively, these operators form a covariance operator field. We show that, in most circumstances, covariance fields are continuous. We also solve the inverse problem: recovering distribution from a covariance field. Surprisingly, this is not possible on Euclidean spaces. On non-Euclidean manifolds however, covariance fields are true distribution representations.

1 Preliminaries

Subject of this study are random variables on Riemannian manifolds. For the sake of clarity and self-consistency we will briefly recall the main notations and facts from Riemannian geometry we are going to use. For a comprehensive introduction the reader is suggested to refer to [4], [11] or [5].

1.1 Riemannian manifolds

Let M be a n -manifold with differentiable structure given as a collection of charts $(U_\alpha, \mathbf{x}_\alpha)$ where U_α are open sets in \mathbb{R}^n and $\mathbf{x}_\alpha : U_\alpha \rightarrow M$ are injective. For $p \in \mathbf{x}_\alpha(U_\alpha)$, $(U_\alpha, \mathbf{x}_\alpha)$ is called a parametrization or system of coordinates

*Florida State University, Department of Statistics

at p . Thus, when we say coordinates x at a point of M , we will understand a local system of coordinates given by a chart (U, \mathbf{x}) .

With M_p we denote the tangent space at $p \in M$. The tangent bundle TM on M is given by $TM = \{(p, v) | p \in M, v \in M_p\}$. It is a $2n$ -manifold. The map $\pi : TM \rightarrow M$, $\pi(p, v) = p$ denotes the natural projection.

Recall that if M_1 and M_2 are two manifolds and $\phi : M_1 \rightarrow M_2$ is a differentiable mapping, differential of ϕ at $p \in M_1$ (also called a push-forward) is a linear mapping $d\phi_p : (M_1)_p \rightarrow (M_2)_{\phi(p)}$ given by $d\phi_p(v)(f) = v(f \circ \phi)$ for any $v \in (M_1)_p$ and $f \in C^\infty(M_2)$, a differentiable function on M_2 . If $d\phi_p$ is an isomorphism, then ϕ is a local diffeomorphism at p (Theorem 2.10 in [4]).

With respect to a parametrization (U, \mathbf{x}) and $p = \mathbf{x}(x_1, \dots, x_n) \in \mathbf{x}(U)$, the tangent space M_p of M at p has canonical basis $\{\frac{\partial}{\partial x_i}|_p\}_{i=1}^n$. Let $v \in M_p$, then $v_x = (v_x^1, \dots, v_x^n) \in \mathbb{R}^n$, such that $v = \sum_{i=1}^n v_x^i \frac{\partial}{\partial x_i}|_p$, is the vector of components of v with respect to coordinates x . A Riemannian structure g on M defines an inner product $\langle \cdot, \cdot \rangle_p$ on M_p such that $g_{ij}(x_1, \dots, x_n) = \langle \frac{\partial}{\partial x_i}|_p, \frac{\partial}{\partial x_j}|_p \rangle_p$ are differentiable functions on U . At each $p \in \mathbf{x}(U)$, the $n \times n$ symmetric and positive definite matrix $G_x = \{g_{ij}(x)\}$ is called a coordinate representation of the metric at p . If y is another local system of coordinates at p and $A = \{\frac{\partial y_j}{\partial x_i}|_p\}_{i=1, j=1}^{n, n}$, is the Jacobian of the change at p , which is a non-singular matrix, then component tangent vectors and metric representations change according to

$$v_y = Av_x, \quad (1)$$

$$G_y = (A^{-1})' G_x A^{-1}. \quad (2)$$

Any Riemannian manifold can be endowed with a natural measure called *volume measure*. Let x be local coordinates at $p \in M$, the volume measure with respect to x is defined by

$$d\mathcal{V}(x) := (d\mathcal{V}(p))_x = \sqrt{\det(G_x)} dx,$$

where dx is the Lebesgue measure in \mathbb{R}^n . One easily checks that a change of local coordinates at p from x to y , does not change the expression for $d\mathcal{V}(p)$, $d\mathcal{V}(x) = d\mathcal{V}(y)$.

Throughout this paper we will assume that M is a Riemannian manifold of dimension n .

1.2 Exponential map and its inverse

Geodesics on M are defined as solutions of first order system of differential equations, called geodesic equations (3.2 in [4]). In local chart (U, \mathbf{x}) they

are

$$\frac{\partial x_k}{\partial t} = y_k, \quad \frac{\partial y_k}{\partial t} = - \sum_{i,j} \Gamma_{ij}^k y_i y_j, \quad (3)$$

where Γ_{ij}^k are differentiable functions in U . The theory of ordinary differential equations says that for any $(x_1, y_1) \in U \times \mathbb{R}^n$, there exists a neighborhood W of (x_1, y_1) and $\epsilon > 0$ such that for any $(x_0, y_0) \in W$, (3) has a unique solution $t \mapsto c(t)$ for $|t| < \epsilon$ satisfying $c(0) = x_0$ and $c'(0) = y_0$. Moreover, $c(t)$ depends differentially on the initial conditions.

For $q \in M$ and $v \in M_q$ let $\gamma(t, q, v)$, $t \in (-\epsilon, \epsilon)$ be a geodesic on M such that $\gamma(0, q, v) = q$ and $\gamma'(0, q, v) = v$. Thus for $x(t) = \mathbf{x}^{-1} \circ \gamma(t, q, v)$ and $y(t) = (\mathbf{x}^{-1} \circ \gamma)'(t, q, v)$, $(x(t), y(t))$ is a solution of the system (3).

For any $p \in M$, there is a set $\mathcal{U} \subset TM$, $p \in \mathcal{U}$ and $\epsilon > 0$, such that $\forall (q, v) \in \mathcal{U}$, $\gamma(t, q, v)$ is well defined and differentiable function of (t, q, v) in $(-\epsilon, \epsilon + 1) \times \mathcal{U}$. Let $\tilde{q} = \gamma(1, q, v)$ and $\tilde{v} = \gamma'(1, q, v)$. Then it follows from (3) that $(\tilde{q}, -\tilde{v}) \in \mathcal{U}$ and for $t \in (-\epsilon, \epsilon + 1)$

$$\gamma(t, \tilde{q}, -\tilde{v}) = \gamma(1 - t, q, v), \quad (4)$$

$$\gamma'(t, \tilde{q}, -\tilde{v}) = -\gamma'(1 - t, q, v).$$

The *exponential map*, $\exp : \mathcal{U} \rightarrow M$, is defined by

$$\exp_q(v) = \exp(q, v) = \gamma(1, q, v).$$

It is a differentiable map on \mathcal{U} .

For any $p \in M$, there is a maximal neighborhood $V(p)$ of the origin in M_p where \exp_p is a diffeomorphism; $U(p) = \exp_p(V(p))$ is called maximal normal neighborhood of p . On $U(p)$, \exp_p has an inverse,

$$\exp_p^{-1} : U(p) \rightarrow V(p) \subset M_p,$$

which is also diffeomorphism.

The differential of the exponential map at $v \in V(p)$

$$(d \exp_p)_v : M_p \rightarrow M_{\exp_p v}$$

is an isomorphism. One checks that $(d \exp_p)_O(v) = \gamma'(0, p, v) = v$ and

$$(d \exp_p)_v(v) = \gamma'(1, p, v). \quad (5)$$

Indeed,

$$(d \exp_p)_v(v) = \frac{d}{dt} \exp_p((t+1)v)|_{t=0} = \frac{d}{dt} \gamma(1, p, (t+1)v)|_{t=0} = \frac{d}{dt} \gamma(t+1, p, v)|_{t=0}.$$

By the Gauss lemma (3.5 in [4]), $(d \exp_p)_v$ also satisfies

$$\langle (d \exp_p)_v(v), (d \exp_p)_v(w) \rangle = \langle v, w \rangle,$$

for any $w \in V(p)$.

Let $p \in U(q)$ and $q \in U(p)$, then by applying (5) and (4) we find

$$\begin{aligned} -\exp_q^{-1} p &= -\gamma'(0, q, \exp_q^{-1} p) = \\ \gamma'(1, p, \exp_p^{-1} q) &= (d \exp_p)_{\exp_p^{-1} q}(\exp_p^{-1} q). \end{aligned}$$

Therefore, for a fixed p , we have following expression of $\exp^{-1} p : U(p) \rightarrow TM$

$$\exp^{-1} p = -\gamma'(1, p, \cdot) \circ \exp_p^{-1}. \quad (6)$$

Since $\gamma'(1, p, \cdot)$ is differentiable in $V(p)$ and \exp_p^{-1} is a diffeomorphism in $U(p)$, $\exp^{-1} p$ is differentiable in $U(p)$.

The map $q \mapsto \exp_q^{-1} p$ is differentiable in $U(p)$ in the following sense. If x are local coordinates at $q \in U(p)$, $q = \mathbf{x}(x_1, \dots, x_n)$, then the components of $(\exp_q^{-1} p)_x \in \mathbb{R}^n$ are differentiable functions of x . Moreover, we show

Lemma 1 *For $q = \mathbf{x}(x_1, \dots, x_n) \in U(p)$, the symmetric matrix*

$$Z_x(q, p) = (\exp_q^{-1} p)_x (\exp_q^{-1} p)'_x \quad (7)$$

is differentiable in x .

A change of local coordinates from x to y with Jacobian A at q , changes coordinate expression of $Z(q, p)$ according to

$$Z_y(q, p) = A Z_x(q, p) A'. \quad (8)$$

We will adopt, for brevity, following notation for the inverse exponential map

$$\vec{qp} := \exp_q^{-1} p,$$

in analogy to the Euclidean case where, $\exp_q^{-1} p = p - q = \vec{qp}$, for $p, q \in \mathbb{R}^n$.

1.3 Tensors and tensor fields

Let V be a n -dimensional vector space and V^* be its *dual* space of linear functions on V . Let also x be a basis of V and \tilde{x} be its dual basis. For $v \in V$ with $v_x \in \mathbb{R}^n$ we denote the column vector of components of v , $v = \sum_i v_x^i x_i$. While for a co-vector $w \in V^*$ with w_x we denote the row vector of components of w , $w'_x \in \mathbb{R}^n$, $w = \sum_i w_x^i \tilde{x}_i$.

Co-variant 2-tensor T on V is a bi-linear function $T : V \times V \rightarrow \mathbb{R}$, which with respect to the basis x is represented by a matrix T_x . Coordinate expression for T is

$$T(u, v) = u'_x T_x v_x, \forall u, v \in V.$$

With $T^2(V)$ we denote the vector space of co-variant 2-tensors on V .

Similarly, contra-variant 2-tensor W is a bi-linear function $W : V^* \times V^* \rightarrow \mathbb{R}$ with a coordinate expression W_x with respect to x

$$W(u, v) = u_x W_x v'_x, \forall u, v \in V^*.$$

With $T_2(V)$ we denote the vector space of contra-variant 2-tensors on V .

Let $T \in T^2(V)$ and $W \in T_2(V)$. The contraction TW of their tensor product $T \otimes W$ is a (1,1) tensor with coordinates expression $T_x W_x$ with respect to x . We denote $TW \in T_1^1(V)$.

Let y be another basis on V such that $y = Ax$ for a non-singular matrix A . Coordinate expressions for T , W and TW change according to

$$T_y = (A^{-1})' T_x A^{-1},$$

$$W_y = A W_x A',$$

and

$$(TW)_y = (A^{-1})' (TW)_x A'.$$

Recall that two matrices C and D are called congruent if there exists a non-singular matrix P such that $C = PDP'$ and are called similar if $C = PDP^{-1}$. Looking back at the change of coordinates rules we may conclude that the coordinate representations of (2,0) and (0,2) tensors are congruent, while those of (1,1) tensors are similar.

Let now M be a Riemmanian manifold with metric structure G . The expression (2) for the change of the metric representation at a point of M is co-variant like. Therefore, at each point $p \in M$ the metric is a symmetric and positive definite co-variant 2-tensor $G(p)$ on the tangent space M_p ; a

fact that we notate with $G(p) \in T^2(M_p)$. Globally, G is a co-variant 2-tensor field which is differentiable in the following sense. If $X, Y \in TM$ are two differentiable vector fields on M , then $G(X, Y)$ is a differentiable function on M . With $\mathcal{T}^2(M)$ ($\mathcal{T}_2(M)$) we denote the differentiable co-variant (contra-variant) 2-tensor fields on M . We write

$$G \in \mathcal{T}^2(M).$$

Now we return back to $Z(q, p)$ which is given by $Z(q, p) = (\overrightarrow{qp})(\overrightarrow{qp})'$, wherever the $\overrightarrow{qp} = \exp_q^{-1} p$ is defined. The change rule (8) for it is a contra-variant like and thus, $Z(q, p)$ is a symmetric and non-negative definite contra-variant 2-tensor at M_q . Moreover, by lemma 1, for any fixed $p \in M$, $Z(., p)$ is a differentiable contra-variant tensor field on $U(p)$ - a fact that we write as

$$Z(., p) \in \mathcal{T}_2(U(p)). \quad (9)$$

Moreover, for every $q \in U(p)$, $G(q)Z(q, p) \in T_1^1(M_q)$ and $G(.)Z(., p)$ is a differentiable (1,1)-tensor field on $U(p)$, i.e.

$$G(.)Z(., p) \in \mathcal{T}_1^1(U(p)). \quad (10)$$

1.4 Linear operators on tangent spaces

Linear operator on vector space V is any $L : V \rightarrow V$ such that

$$L(\alpha v_1 + \beta v_2) = \alpha L(v_1) + \beta L(v_2),$$

for any two $v_1, v_2 \in V$ and $\alpha, \beta \in \mathbb{R}$. With respect to a basis x , L is represented by a matrix L_x and then $L(v) = L_x v_x$. Let y be another basis such that $y = Ax$, then

$$L(v) = L_y v_y = L_y A v_x = (L(v))_y = A(L(v))_x = A L_x v_x.$$

Therefore

$$L_y = A L_x A^{-1},$$

which correspond to the change of coordinates rule for (1,1)-tensors.

Return back to Riemannian manifold setting. Let $p \in M$ and $q \in U(p)$, then $G(q)Z(q, p)$ defines a linear operator on M_q . In local coordinates x at q it is defined as

$$v_x \mapsto (G_x Z_x)' v_x = Z_x G_x v_x, \quad v \in M_q.$$

If $v, w \in M_q$

$$\langle w, (GZ)(v) \rangle = w'_x G_x Z_x G_x v_x.$$

In particular,

$$\langle v, (GZ)(v) \rangle = v'_x G_x Z_x G_x v_x = ((\exp_q^{-1} p)'_x G_x v_x)'((\exp_q^{-1} p)'_x G_x v_x) > 0,$$

for $v \neq 0$.

We summarize in the following

Lemma 2 *For any fixed $p \in M$, $G(\cdot)Z(\cdot, p)$ is differentiable field of linear operators on $U(p)$ and thus, if X is differentiable vector field on $U(p)$, then $(G(\cdot)Z(\cdot, p))(X)$ is also a differentiable vector field on $U(p)$.*

1.5 Distributions on Riemannian manifolds

Let M be a Riemannian manifold and $(U_\alpha, \mathbf{x}_\alpha)$ are charts of M . Open sets $\mathbf{x}_\alpha(U_\alpha)$ in M generate a σ -algebra on M which we will denote with $\mathcal{A}(M)$. One easily verifies that the volume measure, \mathcal{V} , is a measure on the σ -algebra $\mathcal{A}(M)$ and thus, $(M, \mathcal{A}(M), \mathcal{V})$ is a measure space.

A random variable X on M is any measurable function from a probability space $(\Omega, \mathcal{B}, \mathcal{P})$ to $(M, \mathcal{A}, \mathcal{V})$. The distribution function F of X is defined as

$$F(A) = \mathcal{P}(X^{-1}(A)), A \in \mathcal{A}(M).$$

F is a countably additive and satisfies $F \geq 0$, $F(\emptyset) = 0$ and $F(M) = 1$.

Definition 1 *A distribution F on M is said to be absolute continuous with respect to the volume measure if $F(A) = \int_A dF(p), \forall A \in \mathcal{A}(M)$, where $dF(p)$ is given by*

$$dF(p) = f(p)dV(p),$$

for a $\mathcal{A}(M)$ -measurable function f . We say that f is density (pdf) of F .

In the above definition, a density function f is measurable in sense that $f^{-1}(B) \in \mathcal{A}(M)$ for every Borel set B in \mathbb{R} . The density f is continuous everywhere on M except eventually a set of volume measure zero.

In this work we will also consider discrete distributions on M . They are not absolute continuous w.r.t. \mathcal{V} and instead of density have probability mass function (pmf).

2 Covariance fields

2.1 Definition

Definition 2 *Covariance field of probability distribution F is a contra-variant positive definite 2-tensor field Σ on M , given by*

$$q \mapsto \Sigma(q) = \int_{U(q)} (\overrightarrow{qp})(\overrightarrow{qp})' dF(p),$$

where $U(q)$ is the maximal normal neighborhood of q .

In the notation of (7), the covariance of F at q is

$$\Sigma(q) = \int_{U(q)} Z(q, p) dF(p) \in T_2(M_q)$$

because $Z(q, p) \in T_2(M_q)$. At this stage, we do not claim that Σ is differentiable not even continuous field on M or on an open subset of M .

In local coordinates x , $Z_x(q, p)$ is a symmetric non-negative matrix and therefore $\Sigma_x(q) = \int_{U(q)} Z_x(q, p) dF(p)$ is symmetric and non-negative definite. In fact, $\Sigma_x(q)$ will be positive definite, except the cases when the support of F in x coordinates is a hyperplane in \mathbb{R}^n . We ignore these cases, which obviously are caused by ill defined distributions, and assume positive definiteness of $\Sigma_x(q)$. Correspondingly, for the contra-variant tensor $\Sigma(q)$, we assume symmetry and positive definiteness. The space of symmetric and positive definite matrices (tensors) we denote with Sym_n^+ .

At point $q \in M$, we consider the product $G\Sigma(q) := G(q)\Sigma(q)$ of $G(q) \in T^2(M_q)$ and $\Sigma(q) \in T_2(M_q)$. It is a linear operator in M_q , i.e.

$$G\Sigma(q) \in T_1^1(M_q).$$

Let G_x and Σ_x are representations of $G(q)$ and $\Sigma(q)$ with respect to coordinates x about q . Then

$$\langle w, (G\Sigma(q))(v) \rangle = w'_x G_x \Sigma_x G_x v_x,$$

for $v, w \in M_q$. Moreover, $(G\Sigma(q))^{-1}$ is also a linear operator in M_q and

$$\langle w, (G\Sigma(q))^{-1}(v) \rangle = w'_x \Sigma_x^{-1} v_x,$$

If $F(U(q)) = 1$, then

$$\begin{aligned} \text{tr}(G(q)\Sigma(q)) &= \int_{U(q)} (\vec{qp})'_x G_x(q) (\vec{qp})_x dF(p) = \\ &= \int_M ||\text{Log}_q p||^2 dF(p), \end{aligned}$$

Finally, for the intrinsic mean μ , which is the Fréchet mean of F on the metric space M equipped with the geodesic distance, we obtain

$$\mu = \text{argmin}_q \text{tr}(G\Sigma(q)),$$

in agreement to a well known fact about mean of distributions in \mathbb{R}^n .

After this illustrating example we are motivated to give a more general definition of expectation.

Definition 3 *Let F be a distribution on M , $q \in M$, $TW \in T_1^1(M_q)$ and h be a linear operator on $T_1^1(M_q)$. Then the expectation of $h(TW)$ is defined to be*

$$E(h(TW)) = \int_M h(TW(p)) dF(p).$$

As shown above, this is applicable for $h(A) = \text{tr}(A)$ and $TW = G(q)Z(q, p)$. Then $G\Sigma(q) = E(\text{tr}(G(q)Z(q, p)))$, provided that $F(U(q)) = 1$.

2.2 Continuity of $G\Sigma$

We say that a series of points q_k on M converges to a point q_0 and denote $q_k \rightarrow q_0$ if $d(q_k, q_0) \rightarrow 0$, where d is the geodesic distance on M .

Proposition 1 *Let F be a distribution on M and for $q_k \in M$, $k = 0, 1, \dots$, we have $q_k \rightarrow q_0$, as $k \rightarrow \infty$, $F(B(q_k)) = 1$ for all k and $\text{tr}(G\Sigma(q_0)) < \infty$. Then $\text{tr}(G\Sigma(q_k)) \rightarrow \text{tr}(G\Sigma(q_0))$.*

Proof. Without lost of generality we may assume that for all q_k , $d(q_0, q_k) \leq r$. Let X be a random variable with distribution F . Define random variables $\xi_k = \text{tr}(G(q_k)Z(q_k, X))$ and $\xi_0 = \text{tr}(G(q_0)Z(q_0, X))$, where $Z(q, p) = (\vec{qp})(\vec{qp})'$. Observe that

$$F(\cup_k (M \setminus B(q_k))) \leq \sum_k (1 - F(B(q_k))) = 0$$

and therefore $F(\cap_k B(q_k)) = 1$. Since G and Z are continuous at q_0 we have $\xi_k \rightarrow \xi_0$, a.e.. For every $p \in B(q_k) \cap B(q_0)$

$$\begin{aligned} \text{tr}(G(q_k)Z(q_k, p)) &= d^2(q_k, p) \leq (d(q_k, q_0) + d(q_0, p))^2 \\ &\leq (r + d(q_0, p))^2 \leq \max\{(r + 1)^2, r^2 + (2r + 1)d^2(q_0, p)\} \end{aligned}$$

and thus

$$0 \leq \xi_k \leq \max\{(r + 1)^2, r^2 + (2r + 1)\xi_0\}, \text{ a.e..}$$

Finally, since $E\xi_0 < \infty$, the dominated convergence theorem gives us $E(\xi_k) \rightarrow E(\xi_0)$, which is exactly the claim. \square

Proposition 2 *Under the conditions of proposition 1, if $v_k \in M_{q_k} \rightarrow v_0 \in M_{q_0}$ and $w_k \in M_{q_k} \rightarrow w_0 \in M_{q_0}$, then*

$$\langle v_k, (G\Sigma(q_k))(w_k) \rangle \rightarrow \langle v_0, (G\Sigma(q_0))(w_0) \rangle$$

Proof. Define random variables

$$\eta_k = \langle v_k, (G(q_k)Z(q_k, X))(v_k) \rangle$$

and

$$\eta_0 = \langle v_0, (G(q_0)Z(q_0, X))(v_0) \rangle.$$

We have $\eta_k \rightarrow \eta_0$, a.e. and

$$0 \leq \eta_k \leq \|v_k\| \text{tr}(G(q_k)Z(q_k, X)) \|w_k\|.$$

As in Proposition 1, η_k are bounded by a random variable with a finite expectation. Therefore, again by dominated convergence theorem, $E(\eta_k) \rightarrow E(\eta_0)$. \square

We say that covariance field Σ is continuous in the sense given by Proposition 2, i.e. $G\Sigma$ is a continuous field of linear operators on tangent spaces of M .

Definition 4 *The covariance field Σ of a probability distribution F is continuous at $q \in M$ if for any two continuous vector fields v and w on M defined in a neighborhood of q , the function $\langle v, (G\Sigma)w \rangle$ is continuous at q .*

Proposition 1 states sufficient conditions for continuity of Σ . We emphasize that continuity of $G\Sigma$ may hold even when the density of F is discontinuous, provided the conditions are met. Moreover, even discrete distributions may have continuous covariance fields.

Proposition 3 *Let F be a distribution on M , $q_0 \in M$ such that $\text{tr}(G\Sigma(q_0)) < \infty$ and in a neighborhood B_0 of q_0 on M , $F(B(q)) = 1$, for all $q \in B_0$. Then the eigenvalues of $G\Sigma$ are continuous functions at q_0 .*

Proof. Let (\mathbf{x}, U) be a parametrization about q_0 , such that $\mathbf{x}(0) = q_0$ and $\mathbf{x}(U) \subset B_0$. Let $\lambda_i(x)$, $i = 1, \dots, n$, be the eigenvalues of $G\Sigma(\mathbf{x}(x))$, $x \in U$. Consider the canonical tangent vectors $v_i(x) = \frac{\partial}{\partial x_i}|_{\mathbf{x}(x)}$, which form continuous vector fields on $\mathbf{x}(U)$. By Proposition 2, for all i and j

$$\langle v_i, (G\Sigma(q))(v_j) \rangle = [G_x \Sigma_x G_x]_{ij}$$

are continuous functions at 0. Since the elements of matrix G_x are continuous at 0, so are those of $G_x \Sigma_x$. Therefore the eigenvalues $\lambda_i(x)$ are continuous at 0. \square

2.3 Extended covariance fields

In the course of our research we will find useful to extend the definition of covariance field to a whole class $\mathcal{COV}(F)$ of contra-variant tensor fields associated with a particular distribution F . Its members are all

$$\Sigma(q; r) = \int_M (\vec{qp})(\vec{qp})' r(\|\vec{qp}\|) dF(p), \quad (11)$$

where $r : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a continuous function. Thus,

$$\mathcal{COV}(F) = \{\Sigma(q; r) | r \in C(\mathbb{R}^+)\}.$$

The role of r -function in (11) is to control the 'amplitude', $\text{tr}(G\Sigma)$, of the covariance field. It has analytical purpose that is important in numerical experiments. Fields with large amplitude are difficult to be analysed numerically and this fact matters when one develops computational algorithms. We will try to illustrate this with an example.

Example 1 *Let $M = \mathbb{S}^2$ with the standard differential and metric structure (see details in Appendix S). For every $p \in M$, $T_p M = \mathbb{R}^2$ and a normal neighborhood of p is $S(p, \pi)$, the image of the circle $C(0, \pi) \subset T_p M$ under the exponential map, i.e. $S(p, \pi) = \text{Exp}_p(C(0, \pi))$.*

Let F be uniform distribution on M and y_1, \dots, y_k are samples drawn from F . What is the sample covariance field of F based on these samples?

First, we consider the generic covariance field

$$\Sigma_1(y_j) = \sum_{i=1}^k (\overrightarrow{y_j y_i})(\overrightarrow{y_j y_i})',$$

For a second one we apply $r(t) = (1 - \frac{\pi}{2|t|})^2$,

$$\Sigma_2(y_j) = \sum_{i=1}^k (\overrightarrow{y_j y_i})(\overrightarrow{y_j y_i})' (1 - \frac{\pi}{2||\overrightarrow{y_j y_i}||})^2.$$

Let (α, t) be polar coordinates on \mathbb{S}^2 at p . Then $dV(\alpha, t) = \sin(t)d\alpha dt$ and the density is a constant, $f(\alpha, t) = 1/(4\pi)$. As usual with G we denote the metric tensor. First, we calculate

$$E(\text{tr}(G(y_j)(\overrightarrow{y_j y_i})(\overrightarrow{y_j y_i})')) = \frac{1}{2} \int_0^\pi t^2 \sin(t) dt = \frac{\pi^2}{2} - 2,$$

and

$$E(\text{tr}(G(y_j)\Sigma_1(y_j))) = (k-1)(\frac{\pi^2}{2} - 2),$$

while,

$$\begin{aligned} E(\text{tr}(G(y_j)(\overrightarrow{y_j y_i})(\overrightarrow{y_j y_i})') (1 - \frac{\pi}{2||\overrightarrow{y_j y_i}||})^2) = \\ \int_0^\pi t^2 \sin(t) (1 - \frac{\pi}{2t})^2 dt = \frac{\pi^2}{4} - 2 \end{aligned}$$

and

$$E(\text{tr}(G(y_j)\Sigma_2(y_j))) = (k-1)(\frac{\pi^2}{4} - 2).$$

Clearly, $E(\text{tr}(G(y_j)\Sigma_1(y_j))) > 6(k-1)E(\text{tr}(G(y_j)\Sigma_2(y_j)))$.

□

Numerical experiments confirm the benefit of applying an amplitude bounding r -function. A typical choice for r , suggested by experiments, is

$$r(t) = (1 - \frac{R}{2||t||})^2 \tag{12}$$

when the distribution has a bounded domain with geodesic radius R . It is a member of the family $\{r(t) = (1 - \frac{a}{||t||})^2, a > 0\}$.

Recall that a geodesic radius of distribution F is the minimal R such that for every $p \in \text{supp}(F)$, $\text{supp}(F) \subset \text{Exp}_p(C(R))$, where $C(R)$ is the ball with radius R in tangent space at p , i.e. $C(R) = \{v \in \mathbb{R}^n \mid |v| \leq R\}$.

Next result provides sharper estimator of parameter a for the above family of r -functions based on the criterion $\text{tr}(G\Sigma)$ to be minimal. But first we need following definition.

For every point $p \in M$, there are *geodesic spherical coordinates* (θ, t) defined by

$$p(\theta, t) = \text{Exp}_p(t\theta), (\theta, t) \in U(p) \subset \mathbb{S}^{n-1} \times [0, \infty),$$

and $\text{Exp}_p(U(p)) = \mathcal{B}(p)$, the maximal normal neighborhood of p . The change of coordinates at p from normal v to spherical (θ, t) is $dv = t^{n-1}d\theta dt$.

A distribution F is said to have a *bounded density* f , $dF(p) = f(p)dV(p)$, if there exists a positive constant C such that for any $p \in M$, if (θ, t) are the spherical coordinates at p , then

$$f(\theta, t)\sqrt{|G(\theta, t)|} \leq C, \forall \theta, t \in U(p).$$

For example, on the unit 2-sphere, \mathbb{S}^2 ,

$$\sqrt{|G(\theta, t)|}d\theta dt = \sin(t)d\theta dt, (\theta, t) \in [0, 2\pi) \times [0, \pi).$$

and any function f on \mathbb{S}^2 such that $C \geq f \geq 0$, after rescaling, defines a distribution with bounded density.

Lemma 3 *Let \mathcal{F} be a family of distributions on M with geodesic radius R and bounded by C density, then for*

$$r(t; a) = \left(1 - \frac{a}{||t||}\right)^2, \quad (13)$$

we have

$$\sup_{F \in \mathcal{F}} \sup_{q \in M} \text{tr}(G(q)\Sigma(q; a)) \leq \mathcal{V}(\mathbb{S}^{n-1})CR\left(\frac{R^2}{3} - aR + a^2\right),$$

where the covariance field $\Sigma(q, a)$ is defined by (11) with $r(t, a)$ and $\mathcal{V}(\mathbb{S}^{n-1})$ is the volume of \mathbb{S}^{n-1} .

Proof. For $q \in M$ and spherical coordinate system $(\theta, t) \in U \subset \mathbb{S}^{n-1} \times [0, R]$ at p we have

$$\begin{aligned} \text{tr}(G(q)\Sigma(q; a)) &= \int_M \|\vec{qp}\|^2 \left(1 - \frac{a}{\|\vec{qp}\|}\right)^2 dF(p) = \\ &= \int_U t^2 \left(1 - \frac{a}{t}\right)^2 f((\theta, t)) \sqrt{|G((\theta, t))|} d\theta dt. \end{aligned}$$

From the bounded density assumption

$$\text{tr}(G(q)\Sigma(q; a)) \leq C \int_U t^2 \left(1 - \frac{a}{t}\right)^2 d\theta dt \leq$$

$$\mathcal{V}(\mathbb{S}^{n-1}) C \int_0^R t^2 \left(1 - \frac{a}{t}\right)^2 dt = \mathcal{V}(\mathbb{S}^{n-1}) C R \left(\frac{R^2}{3} - aR + a^2\right).$$

Moreover the minimum of the right hand side function of a is achieved for $R/2$. \square

As a consequence of lemma 3, on \mathbb{S}^2 , since all distributions have bounded geodesic radius $R = \pi$, $a = \pi/2$ is the optimal choice for the parameter in (13) and thus, $r(t) = (1 - \frac{\pi}{2\|t\|})^2$ is the optimal member of the family (13).

3 Recovering distribution from covariance

In this section we consider the problem of recovering a distribution from its covariance field. If such a recovery is not possible, then the covariance structure will not be a complete distribution representation and its application potential will be diminished. Fortunately, in most circumstances, the answer of the problem is positive.

3.1 Similarity invariants

The space Sym_n^+ of symmetric and positive definite $n \times n$ matrices is a well studied manifold that accepts a Riemannian structure that makes it a symmetric space. For example, see A. Ohara, N. Suda, S. Amari (1996) and W. Forstner, B. Moonen (1999). We define an important class of functions on Sym_n^+ that respects similarity operation and thus, are functions that can be applied on linear operators.

Definition 5 *A similarity invariant function on Sym_n^+ is any continuous h that satisfies*

(i) $h(AXA', AY A') = h(X, Y)$, $\forall X, Y \in \text{Sym}_n^+$ and $A \in GL_n$.

It is a non-negative with a unique root if

(ii) $h(X, Y) \geq 0$, $\forall X, Y \in \text{Sym}_n^+$ and $h(X, Y) = 0 \iff X = Y$.

Moreover, h is called similarity invariant distance, if in addition to (i) and (ii) also satisfies

(iii) $h(X, Y) + h(Y, Z) \geq h(X, Z)$, $\forall X, Y, Z \in \text{Sym}_n^+$.

We will denote with \mathcal{SIM}_n the class of functions satisfying (i) and (ii).

Below we list several examples of similarity invariant functions.

1. For a fixed $Z \in \text{Sym}_n^+$, the similarity invariant

$$h_{trdif}(X, Y; Z) = |(tr(Z^{-1}X - Z^{-1}Y))|,$$

satisfies (iii) but not (ii). Default choice will be $Z = G^{-1}$, the inverse of the metric tensor representation.

2. The second one is sometimes referred as *affine-invariant distance* in Sym_n^+ , see for example [3], [7] and [14], and it is defined by

$$h_{trln2}(X, Y) = \{tr(\ln^2(XY^{-1}))\}^{1/2}, X, Y \in \text{Sym}_n^+.$$

Actually, h_{trln2} is not a unique choice for a distance in Sym_n^+ .

3. Log-likelihood function gives us another choice for h ,

$$h_{lik}(X, Y) = tr(XY^{-1}) - \ln|XY^{-1}| - n.$$

It satisfies (i) and (ii) but it fails to satisfy the triangular inequality.

- 4.

$$h_{lntr}(X, Y) = \ln(tr(XY^{-1} - YX^{-1})^2),$$

h_{lntr} is another candidate for a distance in Sym_n^+ .

5. Another interesting choice for h is

$$h_{lnpr}(X, Y) = \ln(tr(XY^{-1})tr(YX^{-1})),$$

that satisfies (iii) and 'almost' satisfies (ii): $h_{lnpr}(X, Y) = 0 \iff X = cY$, for $c > 0$.

Let F be a distribution on M and Σ be its covariance field. Consider a similarity invariant h applied on covariance operator field $G\Sigma$. If the conditions of Proposition 3 hold, $h(G\Sigma)$ would be a continuous scalar field on M . Indeed, h is a continuous function of the eigenvalues of $G\Sigma$, which are continuous by themselves.

Scalar fields of the form $h(G\Sigma)$ can be viewed as *representations* of F . For some choices of h , they would be true distribution representations, in sense that the underlying distribution can be fully recovered from them. And this is what we are going to address next.

3.2 Recovering discrete distributions

Let $\mathcal{P} = \{p_i\}_{i=1}^k$ be a set of k points on M . By a discrete mass function (*pmf*) on M we understand any f defined on the domain set \mathcal{P} , such that $f = \{f_i = f(p_i) \geq 0\}_{i=1}^k$ and $\sum_{i=1}^k f_i = 1$. We write $f \in P_k^+$, where P_k^+ denotes the compact k -simplex.

Let $\mathcal{Q} = \{q_j\}_{j=1}^k$ be another set of k points on M , called *observation set*, where the covariances of $f \in P_k^+$ will be considered.

We assume that the set \mathcal{P} is contained within the maximal normal neighborhood of each of the points q_j in order for the vectors $\overrightarrow{q_j p_i}$ to be well defined. This assumption is not a strong one when M is a complete Riemannian manifold. Thus, for every j we may assume a fixed local parametrization (x_j, U_j) such that $\mathcal{P} \subset x_j(U_j)$.

Fix an amplitude controlling function r . Covariance of $f \in P_k^+$ at q_j is defined as

$$\Sigma[f]_j := \Sigma[f](q_j) = \sum_{i=1}^k (\overrightarrow{q_j p_i})(\overrightarrow{q_j p_i})' r(\|\overrightarrow{q_j p_i}\|) f_i.$$

Let us denote

$$Y_{ji} = (\overrightarrow{q_j p_i})(\overrightarrow{q_j p_i})' r(\|\overrightarrow{q_j p_i}\|), i = 1 \dots k.$$

then $\Sigma[f]_j = \sum_{i=1}^k f_i Y_{ji}$. The collection $\{\Sigma[f]_j\}_{j=1}^k$ is called a covariance set of f on \mathcal{Q} .

Now we are interested in the inverse problem. How to reconstruct a *pmf* from its observed covariances? Let $\mathcal{C} = \{C_j \in T_2(M_{q_j})\}_{j=1}^k$ be a set of contravariant tensors, which may happen to be covariance tensors of an unknown distribution or may not. The problem is to find f such that

$$\Sigma[f]_j \approx C_j, j = 1 \dots, k.$$

To measure the 'closeness' we will use similarity invariant.

Define the functional

$$H(f) = \frac{1}{k} \sum_{j=1}^k h(\Sigma[f]_j, C_j), f \in P_k^+, \quad (14)$$

where $h \in \mathcal{SIM}(n)$. Now we can formulate more precisely our problem as an optimization one: find a $pmf \hat{f}$ such that

$$\hat{f} = \operatorname{argmin}_f H(f). \quad (15)$$

From the assumptions for h , it is clear that

$$H(f) = 0 \iff h(\Sigma[f]_j, C_j) = 0, \forall j \iff \Sigma[f]_j = C_j, \forall j.$$

If C_j are covariances that come from a $pmf f^0 \in P_k^+$, then f^0 will be a solution of the system

$$\sum_{i=1}^k f_i Y_{ji} = C_j, j = 1 \dots k. \quad (16)$$

To be able to recover correctly f^0 , the system (16) should have a unique solution. A necessary and sufficient condition for that is

$$\operatorname{rank}(\mathcal{Y}|\mathcal{C}) = \operatorname{rank}(\mathcal{Y}) = k,$$

where $\mathcal{Y} = \mathcal{Y}[q_j] := \{d_{ji} = \operatorname{tr}(G(q_j)Y_{ji})\}_{j=1, i=1}^{k,k}$, $\mathcal{C} = \{\operatorname{tr}(G(q_j)C_j)\}_{j=1}^k$ and $\mathcal{Y}|\mathcal{C}$ is the matrix \mathcal{Y} with vector \mathcal{C} attached as a last column. Note that $d_{ji} = d^2(q_j, p_i)r(d(q_j, p_i))$ and for $r = 1$ these are the squared geodesic distances.

Definition 6 *We say that a covariance operator field $G\Sigma$ on M has a full rank, if for any bounded subset $A \subset M$, $k \in \mathbb{N}$ and k -sample $\{q_j\}_{j=1}^k$, selected by a continuous distribution Q on A , we have $P_Q(\operatorname{rank}(\mathcal{Y}) = k) = 1$, where $\mathcal{Y} = \{d^2(q_j, p_i)r(d(q_j, p_i))\}_{i=1, j=1}^{k,k}$.*

In Eulclidean space, $M \equiv \mathbb{R}^n$, the rank of $\{d^2(q_j, p_i)\}_{j=1, i=1}^{k,k}$ is bounded above by $n + 2$, i.e. for a default covariance field in Euclidean space,

$$\operatorname{rank}(\mathcal{Y}) \leq n + 2,$$

and (16) does not have a unique solution when $k > n + 2$. The problem can be fixed using a non-default covariance field with amplitude function $r \neq 1$ in (11).

We hypothesize that on non-Euclidean space, a manifold with non-zero curvature, any covariance operator field is of full rank. Experiments on spheres, manifolds with constant sectional curvature +1, and hyperbolic plane, a manifold with constant sectional curvature -1, confirm the hypothesis, but of course a more formal argument is needed here.

If matrix \mathcal{Y} has full rank, then one can find the pmf f directly solving system (16) or minimizing the functional $H(f)$ as given in (14) for $h \in \mathcal{SIM}_n$. The second choice is much more general and gives us solutions even in cases when C_j are not in fact true covariances, i.e. $rank(\mathcal{Y}|\mathcal{C}) > rank(\mathcal{Y})$.

It is also important to know in what cases the optimization problem (15) can be solved easily. A function h for which the corresponding functional H is convex is an obvious choice since in that case it is straightforward to find the global minimum of H by the gradient descend algorithm.

Definition 7 *We say that $h \in \mathcal{SIM}_n$ is convex (in Sym_n^+) if for any k and $Y_{ji}, C_j \in Sym_n^+$, $i, j=1, \dots, k$, such that $rank(\mathcal{Y}) = k$, $\sum_{j=1}^k h(\sum_{i=1}^k f_i Y_{ji}, C_j)$ is a convex function of f in P_k^+ .*

From the list of invariants we list above, h_{trdif}^2 , h_{lik} and $h_{trs q}^2$ are convex. We will show it for the last one.

Example 2 *Let Y_{ji} be such that $rank(\mathcal{Y} = \{tr(Y_{ji})\}_{j,i}) = k$. We will show that for $h_{trs q}^2(A, B) := tr((AB^{-1} - BA^{-1})^2)$, the functional*

$$H(f) = \sum_{j=1}^k tr((\sum_{i=1}^k f_i Y_{ij}) - (\sum_{i=1}^k f_i Y_{ij})^{-1})^2$$

is convex. Without loss of generality we assume $C_i = I_n$. Then

$$\frac{\partial H}{\partial f_s} = 2 \sum_{j=1}^k tr[Y_{js}(\sum_{i=1}^k f_i Y_{ij}) - (\sum_{i=1}^k f_i Y_{ij})^{-1} Y_{sj}(\sum_{i=1}^k f_i Y_{ij})^{-2}]$$

and defining $B_j = \sum_{i=1}^k f_i Y_{ij}$, we obtain

$$\frac{\partial H}{\partial f_s} = 2 \sum_{j=1}^k tr[Y_{js}(B_j - B_j^{-3})].$$

Second derivatives are

$$\frac{\partial^2 H}{\partial f_s \partial f_l} = 2 \sum_j \text{tr} \{ Y_{sj} Y_{lj} + Y_{sj} B_j^{-1} Y_{lj} B_j^{-3} + Y_{sj} B_j^{-2} Y_{lj} B_j^{-2} + Y_{sj} B_j^{-3} Y_{lj} B_j^{-1} \}.$$

For $w \in \mathbb{R}^k$ and $w \neq 0$, let $Z_j = \sum_{s=1}^k z_s Y_{sj}$, then

$$w' \frac{\partial^2 H}{\partial f^2} w = 2 \sum_{j=1}^k \text{tr} \{ Z_j Z_j + Z_j B_j^{-1} Z_j B_j^{-3} + Z_j B_j^{-2} Z_j B_j^{-2} + Z_j B_j^{-3} Z_j B_j^{-1} \}.$$

Since $\text{rank}(\mathcal{Y}) = k$, $\sum_{j=1}^k \text{tr}(Z_j Z_j) > 0$. Also,

$$\text{tr}(Z_j B_j^{-1} Z_j B_j^{-3}) = \text{tr}([B_j^{-2} Z_j B_j^{-1}][B_j^{-1} Z_j B_j^{-2}]') \geq 0$$

and similarly $\text{tr}(Z_j B_j^{-2} Z_j B_j^{-2}) \geq 0$. That proves the convexity of h_{trsq}^2 . \square

Now we are interested in the problem of consistency of estimators (15). If we assume that the covariances C_j are random and converge in probability to some matrices, is it true that the estimators \hat{f} also converge? To guarantee a positive answer we need more assumptions for the invariant h .

Theorem 1 Let $h \in \mathcal{SIM}_n$ be a distance function. Let also $f^0 \in P_k^+$ be a pmf on \mathcal{P} and $\{C_j^0 = \Sigma[f^0](q_j)\}_{j=1}^k$ be its covariance set on \mathcal{Q} . If $\mathcal{C}^m = \{C_j^m \in \text{Sym}_n^+\}_{j=1}^k$ is a sequence of set of random matrices such that

$$\forall j = 1, \dots, k, \quad h(C_j^m, C_j^0) \xrightarrow{p} 0, \quad \text{as } m \rightarrow \infty$$

then for

$$\hat{f}^m = \text{argmin}_{f \in P_k^+} H_m(f), \quad \text{where } H_m(f) = \frac{1}{k} \sum_{j=1}^k h(\Sigma[f]_j, C_j^m),$$

we have $\hat{f}^m \xrightarrow{p} f^0$.

Proof.

Since h satisfies the triangular inequality, then for any $f \in P_k^+$

$$|h(\Sigma[f]_j, C_j^m) - h(\Sigma[f]_j, C_j^0)| \leq h(C_j^m, C_j^0)$$

and

$$\sup_{f \in P_k^+} |h(\Sigma[f]_j, C_j^m) - h(\Sigma[f]_j, C_j^0)| \leq h(C_j^m, C_j^0).$$

By summing on j we obtain

$$\sup_{f \in P_k^+} |H_m(f) - H_0(f)| \leq \sum_{j=1}^k h(C_j^m, C_j^0).$$

Under the assumptions on h , $H_0(f)$ has a well-separated minimum at $f^0 \in P_k^+$. In fact $H_0(f^0) = 0$. Therefore, for any $\delta > 0$, there exists $\epsilon > 0$, such that

$$H_0(f) > H_0(f^0) + 2\epsilon, \text{ for any } f \text{ such that } |f - f^0|_{L_2} > \delta.$$

Then we have

$$P(|\hat{f}^m - f^0|_{L_2} > \delta) \leq P(H_0(\hat{f}^m) > H_0(f^0) + 2\epsilon) \leq$$

$$P(H_0(\hat{f}^m) - H_m(\hat{f}^m) + H_m(\hat{f}^m) - H_0(f^0) > 2\epsilon) \leq$$

since $H_m(\hat{f}^m) \leq H_m(f^0)$

$$P(H_0(\hat{f}^m) - H_m(\hat{f}^m) + H_m(f^0) - H_0(f^0) > 2\epsilon) \leq$$

$$P(2\sup_{f \in P_k^+} |H_m(f) - H_0(f)| > 2\epsilon) \leq$$

$$P\left(\sum_{j=1}^k h(C_j^m, C_j^0) > k\epsilon\right) \leq \sum_{j=1}^k P(h(C_j^m, C_j^0) > \epsilon).$$

Since for any j , $P(h(C_j^m, C_j^0) > \epsilon) \longrightarrow 0$, we have $P(|\hat{f}^m - f^0|_{L_2} > \delta) \longrightarrow 0$.

□

Unfortunately, the above result is not very useful in practice for distance functions are usually non-convex and non-convexity of H makes its optimization difficult. In fact, the condition on h to satisfy the triangular inequality is a stronger assumption than what we actually need. We observe that it is only used to bound uniformly $|h(\Sigma[f]_j, C_j^m) - h(\Sigma[f]_j, C_j^0)|$ by $h(C_j^m, C_j^0)$. Therefore, if we guarantee the uniform convergence of the former difference by other means, the triangular inequality condition will be redundant.

Consistency Criterion 1 *We say that a similarity invariant function h satisfies the consistency criterion, if for $C^m \in \text{Sym}_n^+$ random, and $B_i, C^0 \in \text{Sym}_n^+$,*

$$h(C^m, C^0) \longrightarrow_p 0$$

implies

$$\sup_f |h(\sum_{i=1}^k f_i B_i, C^m) - h(\sum_{i=1}^k f_i B_i, C^0)| \longrightarrow_p 0, \text{ as } m \rightarrow \infty.$$

The following theorem is a stronger version of Theorem 1, but using the above consistency criterion, and can be proven similarly.

Theorem 2 *Let $h \in \mathcal{SIM}_n$ satisfy consistency criterion 1. Let also $f^0 \in P_k^+$ be a pmf on \mathcal{P} and $\{C_j^0 = \Sigma[f^0](q_j)\}_{j=1}^k$ be its covariance set on \mathcal{Q} . If $C^m = \{C_j^m \in \text{Sym}_n^+\}_{j=1}^k$ is a sequence of set of random matrices such that*

$$\forall j = 1, \dots, k, \quad h(C_j^m, C_j^0) \longrightarrow_p 0, \text{ as } m \rightarrow \infty$$

then $\hat{f}^m \longrightarrow_p f^0$.

It turns out that $h_{\text{trln}2}$ invariant is a distance but is not convex in the sense of definition (7) and finding the global minimum of $H_{\text{trln}2}$ is difficult. On the other hand invariants h_{lik} and h_{trsq}^2 are both convex and satisfy the consistency criterion 1, which makes them better choices.

The condition $h(C_j^m, C_j^0) \longrightarrow_p 0$ can be further simplified if h is continuous. We say that a sequence X_m of random $n \times n$ matrices converges in probability to matrix C and write $X_m \longrightarrow_p C$, if for any $v \in \mathbb{R}^n$, $v'(X_m - C)v \rightarrow_p 0$. One can easily check that if $h(X, C)$ is continuous in $X \in \text{Sym}_n^+$ for every $C \in \text{Sym}_n^+$, then $X_m \longrightarrow_p C$ if and only if $h(X_m, C) \rightarrow_p 0$. Next we have the following corollary of Theorem 2.

Corollary 1 *Let $h \in \mathcal{SIM}_n$ be a continuous invariant that satisfies consistency criterion 1. If $C_j^m \longrightarrow_p C_j^0$, $\forall j = 1, \dots, k$, then $\hat{f}^m \longrightarrow_p f^0$.*

Note that both h_{lik} and h_{trsq}^2 satisfy the conditions of the above corollary.

3.3 Recovering continuous distributions

We will give a constructive procedure for recovering a continuous distribution density from its covariance operator field. It turns out that any covariance field of full rank specifies completely the underlying density when its domain is a bounded compact. This fact shows that covariance fields are in general faithful representations of corresponding distributions. Although this fact may seem obvious at first look, there is a notable exception that makes the problem of recovering relevant. In Euclidean space, $M = \mathbb{R}^n$, we have the following relation for the default covariance field of random variable X with $E(X) = \mu$

$$E[(q-x)(q-x)'] = E[(p-x)(p-x)'] + (q-p)(q-p)' + (q-p)(p-\mu)' + (p-\mu)(q-p)',$$

Thus

$$\Sigma(q) = \Sigma(\mu) + (q - \mu)(q - \mu)'$$

does not contain any information beyond the first two moments: the mean μ and the covariance $\Sigma(\mu)$. Therefore the covariance field is defined only by the first two moments of X and can not possibly represent the whole distribution.

We use the same approach as for recovering discrete distributions and start with selecting appropriate similarity invariants as technical instruments. We need to make stronger assumptions than that in consistency criterion (1).

Consistency Criterion 2 *We say that similarity invariant function h satisfies the consistency criterion if for $B_i, C_1, C_2 \in \text{Sym}_n^+$ such that $\|B_i\| \leq \gamma$ and $\|C_i\| \leq \gamma$*

$$\sup_f |h(\sum_{i=1}^k f_i B_i, C_1) - h(\sum_{i=1}^k f_i B_i, C_2)| \leq \alpha \gamma \|C_1 - C_2\|,$$

for a constant $\alpha > 0$, independent of B_i and C_i .

Example 3 *We will show that $h_{\text{trdif}}^2(A, B) = \text{tr}^2(A - B)$ is convex and satisfies the consistency criterion 2. Let Y_{ji} are such that $\text{rank}(\mathcal{Y} = \{\text{tr}(Y_{ji})\}_{j,i}) = k$. Define $L(f) = \sum_{j=1}^k (\sum_{i=1}^k f_i \text{tr}(Y_{ij}) - \text{tr}(C_j))^2$, then*

$$\frac{\partial L}{\partial f_s} = 2 \sum_{j=1}^k \text{tr}(Y_{js}) (\sum_{i=1}^k f_i \text{tr}(Y_{ij}) - \text{tr}(C_j))$$

and

$$\frac{\partial^2 L}{\partial f_s \partial f_l} = 2 \sum_{j=1}^k \text{tr}(Y_{js}) \text{tr}(Y_{jl})$$

For $w \in \mathbb{R}^k$,

$$w' \frac{\partial^2 L}{\partial f^2} w = 2 \sum_{j=1}^k \left(\sum_{s=1}^k w_s \text{tr}(Y_{js}) \right)^2 > 0$$

because $\text{rank}(\mathcal{Y}) = k$. This shows the convexity.

Next, observe that for B_i , C_1 and C_2 , such that $\|B_i\| \leq \gamma$, $\|C_i\| \leq \gamma$

$$\left| \left(\sum_{i=1}^k f_i \text{tr}(B_i) - \text{tr}(C_1) \right)^2 - \left(\sum_{i=1}^k f_i \text{tr}(B_i) - \text{tr}(C_2) \right)^2 \right| \leq$$

$$|\text{tr}(C_1) - \text{tr}(C_2)| \left| 2 \text{tr} \left(\sum_{i=1}^k f_i B_i \right) + \text{tr}(C_1) + \text{tr}(C_2) \right| \leq 2n(n+1)\gamma \|C_1 - C_2\|,$$

since for any $n \times n$ matrix X , $|\text{tr}(X)| \leq n\|X\|$. \square

In order to show our main result we need to put some restrictions on the size of distribution domains. In the sequel, d_g denotes the geodesic distance corresponding to a metric g on M .

Definition 8 We say that the inverse exponential map \exp^{-1} on M is Lipschitz in $A \subset M$, if there exists a constant $\beta > 0$ such that $\forall q \in A$

$$\|\overrightarrow{qp_1} - \overrightarrow{qp_2}\|_{L^2} = \|\exp_q^{-1} p_1 - \exp_q^{-1} p_2\|_{L^2} \leq \beta d_g(p_1, p_2), \forall p_1, p_2 \in A.$$

Necessary condition for \exp^{-1} to be Lipschitz in A is for any $q \in A$, $A \subset U(q)$, where, recall, $U(q)$ denotes the domain where \exp_q^{-1} is defined and is in fact diffeomorphic. It is an open question whether this condition is sufficient one. The next example provides some evidence in its support.

Example 4 On the unit 2-sphere, $M = \mathbb{S}^2$, the \exp^{-1} -map is Lipschitz in any compact K with geodesic diameter less than π . If $\rho = \text{diam}(K) < \pi$ then parameter $\beta = \frac{\rho}{\sin(\rho)}$.

Theorem 3 *Let F be a distribution on M and let K be a compact in M , which is bounded, $\text{diam}(K) \leq R$, contains the support of F , $F(K) = 1$, and Log-map is Lipschitz in K . Let also the default covariance field Σ of F has a full rank a.e. on M . Then there is a sequence \hat{F}^m of pmfs on M obtained from the field $G\Sigma$ alone, such that $\forall V \subset M$*

$$\hat{F}^m(V) \longrightarrow F(V), \text{ as } m \rightarrow \infty.$$

Proof. Under the assumptions for K , for any $m > 0$ there exists a partition $\{U_j^m\}_{j=1}^{N(m)}$ of K with geodesic diameter no greater than $1/m$ and each point $q_j^m \in U_j^m$ is selected independently by the uniform distribution on U_j^m . Then $P(\text{rank}(\mathcal{Y}[q_j^m]) = N(m)) = 1$.

Define $\tilde{f}_j^m = F(U_j^m)$ and

$$\tilde{\Sigma}^m(q) = \sum_{j=1}^{N(m)} \tilde{f}_j^m (\overrightarrow{qq_j^m})(\overrightarrow{qq_j^m})', q \in K.$$

Note that $\tilde{\Sigma}^m(q)$ is random because it depends on the choice of q_j^m . Since

$$\begin{aligned} & \sum_{j=1}^{N(m)} (\overrightarrow{qq_j^m})(\overrightarrow{qq_j^m})' - \int_K (\overrightarrow{qp})(\overrightarrow{qp})' dF(p) = \\ & \sum_{j=1}^{N(m)} \int_{U_j^m} [(\overrightarrow{qq_j^m})(\overrightarrow{qq_j^m})' - (\overrightarrow{qp})(\overrightarrow{qp})'] dF(p), \end{aligned}$$

for any $v \in M_q$ we have

$$\begin{aligned} & | \langle v, (G\tilde{\Sigma}^m(q) - G\Sigma(q))(v) \rangle | = \\ & \sum_{j=1}^{N(m)} \int_{U_j^m} | \langle v, \overrightarrow{qq_j^m} \rangle^2 - \langle v, \overrightarrow{qp} \rangle^2 | dF(p) \leq \\ & \sum_{j=1}^{N(m)} \int_{U_j^m} \|v\| \|\overrightarrow{qq_j^m} - \overrightarrow{qp}\| \|v\| (\|\overrightarrow{qq_j^m}\| + \|\overrightarrow{qp}\|) dF(p) < \\ & \|v\|^2 \beta (2R^2 + \frac{1}{m}) \frac{1}{m}, \text{ w. p. } 1. \end{aligned}$$

Thus the operator norm of $G\tilde{\Sigma}^m(q) - G\Sigma(q)$ is uniformly bounded

$$||G\tilde{\Sigma}^m(q) - G\Sigma(q)|| \leq \beta(2R^2 + \frac{1}{m})\frac{1}{m}, \text{ w. p. } 1. \quad (17)$$

In particular, since $\dim(M)=n$

$$\max_{q \in A} |tr(G\tilde{\Sigma}^m(q) - G\Sigma(q))| \leq n\beta(2R^2 + \frac{1}{m})\frac{1}{m}, \text{ w. p. } 1.$$

Since Σ has a full rank *a.e.* on M , then there exists a $h \in \mathcal{S}\mathcal{I}\mathcal{M}_n$ that is convex (in sense of definition 7) and satisfies consistency criteria 2 (w.p. 1). For example, we may take $h = h_{trdif}^2$. Define

$$H_m(f) = \frac{1}{N(m)} \sum_{j=1}^{N(m)} h(G\Sigma[f](q_j^m), G\Sigma(q_j^m))$$

$$\tilde{H}_m(f) = \frac{1}{N(m)} \sum_{j=1}^{N(m)} h(G\Sigma[f](q_j^m), G\tilde{\Sigma}^m(q_j^m))$$

and let

$$\hat{f}^m = \operatorname{argmin}_f H_m(f) \text{ and } \tilde{f}^m = \operatorname{argmin}_f \tilde{H}_m(f).$$

Since $\tilde{H}_m(\tilde{f}^m) = 0$, $\tilde{H}_m(f)$ has a well separated minimum at \tilde{f}^m , i.e. for any $\delta > 0$, there exists $\epsilon > 0$, such that

$$\tilde{H}_m(f) > \tilde{H}_m(\tilde{f}^m) + 2\epsilon = 2\epsilon, \text{ for any } f \text{ such that } |f - \tilde{f}^m|_{L_2} > \delta.$$

Therefore

$$P(|\hat{f}^m - \tilde{f}^m|_{L_2} > \delta) \leq P(\tilde{H}_m(\hat{f}^m) > \tilde{H}_m(\tilde{f}^m) + 2\epsilon) =$$

$$P(\tilde{H}_m(\hat{f}^m) - H_m(\hat{f}^m) + H_m(\hat{f}^m) - \tilde{H}_m(\tilde{f}^m) > 2\epsilon) \leq$$

since $H_m(\hat{f}^m) \leq H_m(\tilde{f}^m)$

$$P(\tilde{H}_m(\hat{f}^m) - H_m(\hat{f}^m) + H_m(\tilde{f}^m) - \tilde{H}_m(\tilde{f}^m) > 2\epsilon) \leq$$

$$P(2 \sup_{f \in P_k^+} |H_m(f) - \tilde{H}_m(f)| > 2\epsilon) \leq$$

by consistency criteria 2 for h and the fact that $\|(\vec{qp})(\vec{qp})'\| \leq nR^2$ on K

$$P(\alpha n R^2 \frac{1}{N(m)} \sum_{j=1}^{N(m)} \|G\tilde{\Sigma}^m(q_j^m) - G\Sigma(q_j^m)\| > \epsilon) \longrightarrow 0, \text{ as } m \rightarrow \infty$$

by the virtue of (17). Thus $\hat{f}^m - \tilde{f}^m \longrightarrow_p 0$.

Finally, for any $V \subset M$ we have

$$\sum_{j:q_j^m \in V} \tilde{f}_j^m \longrightarrow_p F(V), \text{ as } m \rightarrow \infty$$

and therefore

$$\hat{F}^m(V) = \sum_{j:q_j^m \in V} \hat{f}_j^m \longrightarrow_p F(V).$$

□

The above theorem is constructive and give us the freedom to choose a similarity invariant h provided it is convex and satisfies the consistency criterion 2. As we showed, h_{trdif}^2 satisfies both requirements and can be applied. In this case the corresponding optimization functional

$$H_m(f) = \frac{1}{N(m)} \sum_{j=1}^{N(m)} \left(\sum_i d^2(q_i^m, q_j^m) - \rho(q_j^m) \right)^2$$

is based on the function $\rho(q) = E_F d^2(q, p) = \int_K d^2(q, p) dF(p)$, the "variance" of F with respect to q . What the theorem states, basically, is that distribution F can be recovered from the scalar field ρ on M provided that the full-rank condition for the covariance field on M holds and in the domain of F , \exp^{-1} -map is Lipschitz.

The requirement for h to satisfy the consistency criterion 2 can be relaxed. We only need $h \in \mathcal{SIM}_n$ to be continuous invariant that satisfies consistency criterion 1. Indeed, looking back at the proof above we see that $\|G\tilde{\Sigma}^m(q_j^m) - G\Sigma(q_j^m)\| \longrightarrow_p 0$, and therefore $\sup_{f \in P_k^+} |H_m(f) - \tilde{H}_m(f)| \longrightarrow_p 0$, which is what we needed to show $\hat{f}^m - \tilde{f}^m \longrightarrow_p 0$. That observation gives us more choices for h such as the convex invariants h_{lik} and h_{trsqr}^2 .

In \mathbb{R}^n recovering from default covariance field is not possible because the full-rank condition fails. On non-Euclidean (with non-zero curvature) spaces however, the full-rank condition is generally true and reconstruction

is possible. For example, if $M = \mathbb{S}^2$ and $\text{supp}(F) \subset K$, for a compact on \mathbb{S}^2 with $\text{diam}(K) < \pi$, then the theorem is applicable and F can surely be recovered.

Some of the conditions of the above theorem can be relaxed. For example, as defined the full-rank condition for the covariance field allow infinitely many choices for the points q_j^m and thus infinitely many choices of sequence \hat{F}^m converging to F . Different invariants h also give different approximating distributions \hat{F}^m .

Theorem 1 shows how to recover discrete distribution provided the full-rank condition only. Lipschitz condition in theorem 3 is a technical one and it might be possible to relax it in a different approach to the problem. Another opportunity, for example, is to work with a covariance field with amplitude $r \neq 1$ and to satisfy both full-rank condition and Lipschitz condition, even in \mathbb{R}^n .

3.4 Summary

We introduced the concept of covariance field of a distribution on a Riemannian manifold. It is a contra-variant 2-tensor field on the manifold. Closely associated with it is a covariance operator field, which defines a linear operator on the tangent space at each point on the manifold. Covariance operator fields, in most cases, are continuous.

Covariance fields, in general, recover the underlying distributions and in this sense are faithful distribution representations. There is one major exception though where such reconstruction is not possible, the Euclidean space \mathbb{R}^n . This interesting fact shows that the covariance field concept is indeed more relevant to non-Euclidean manifolds.

References

- [1] Amari, S.,-I. (1985) Differential-Geometrical Methods in Statistics. *Lecture Notes in Statistics*, Springer, New York.
- [2] Bhattacharya R. and Patrangenaru, V. (2003) Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds - I *The Annals of Statistics*, Vol.31(1), 1-29.

- [3] Batchelor, P., Moakher, M., Atkinson, D., Calamante, F., Connelly, A. (2005) A rigorous framework for diffusion tensor calculus. *Magn. Reson. Med.*, 53(1), 221-225.
- [4] M.P. Do Carmo. *Riemannian Geometry*, Birkhauser, Boston, 1992.
- [5] I. Chavel. *Riemannian Geometry: A Modern Introduction*, Cambridge University Press, 1993.
- [6] Dryden, I., Koloydenko, A., and Zhou, D., (2008) Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. University of Nottingham, NG7 2RD, UK.
- [7] Fletcher, P. T., Joshi, S., (2007) Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250-262.
- [8] W. Forstner, B. Moonen. (1999) A Metric for Covariance Matrices. Stuttgart University.
- [9] Helgason, S., (1978) *Differential Geometry, Lie Groups, and Symmetric Spaces*. *Academic Press*. New York
- [10] H. Karcher. (1977) Riemannian center of mass and mollifier smoothing, *Communications on Pure and Applied Mathematics*, 30(5):504-541.
- [11] John M. Lee. *Introduction to Smooth manifolds*, 2000.
- [12] Muchor, P., Mumford, D., 2006 Riemannian geometries on spaces of plane curves. *J. of the European Math. Soc.* 8, 1-48.
- [13] A. Ohara, N. Suda, S. Amari. Dualistic Differential Geometry of positive definite matrices and its applications to related problems. *Linear Algebra and Its Applications* 247:31:53, 1996.
- [14] Pennec. X., Fillard, P., Ayache, N (2006), A Riemannian framework for tensor computing. *Int. J. Computational Vision*, 66(1):41-66.